On-Demand Scientific Computing

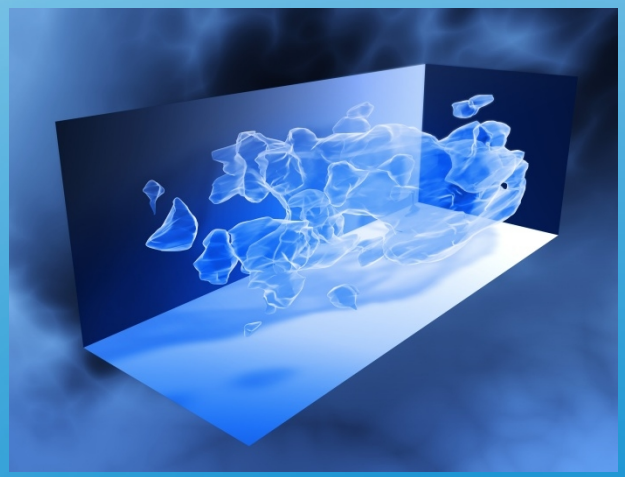
LA-UR 10-05199

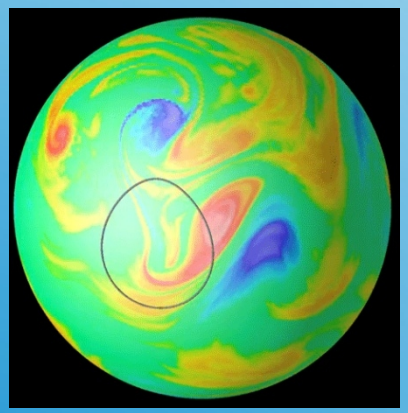
Travis Johnson | Gregorio Hinojos | Benjamin Cheng

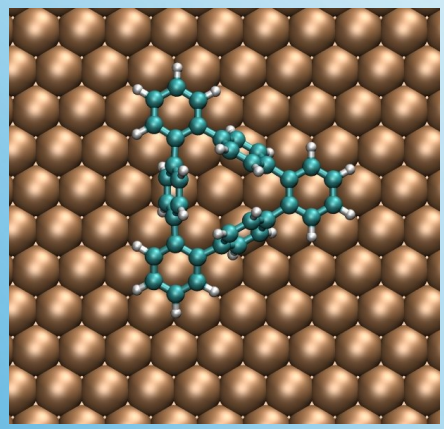
On-Demand…

NETFLIX

Scientific Computing







Science on traditional HPC systems

The Problem:

- For the scientist that wants to run a large job, they may have to wait for other jobs to finish.
- As systems upgrade, older software cannot run on the new systems
- Jobs cannot be set to run at the discretion of it's user.
- Clusters are expensive to build and maintain

How does virtualization fit in to help solve this problem?

No Virtualization for HPC…

Virtualization has not been used in traditional scientific clusters

Virtualization conflicts with the primary goal of solving larger problems faster.

So How do we add virtualization to the mix?

What we know of virtualization

- Virtualization allows consolidation of various supporting tasks
- Virtualization can flexibly allocate resources.
- Virtualization can support most any OS's old and new

To fit virtualization, we are taking what we know about virtualization and creating an alternative

Providing an alternative for Scientists

Our goal is not to replace the current HPC structure, but to provide a supplementary virtualized system that Scientists can run jobs on.

Virtualization + On-demand = Infrastructure On-Demand

Infrastructure On-Demand = Better Utilization of HPC

Science on traditional HPC Systems with the option of Infrastructure On-Demand

The Solution:
Place small jobs somewhere else

- Reduces the load for the HPC
- Makes things more efficient
- Produces a more reliable and enjoyable HPC experience.

Better HPC utilization is the goal!

Creating a Scientific Virtual Appliance

Travis Johnson | Gregorio Hinojos | Benjamin Cheng

Institute Bootcamp

ISTI
INFORMATION SCIENCE & TECHNOLOGY INSTITUTE

- Creating a Cluster

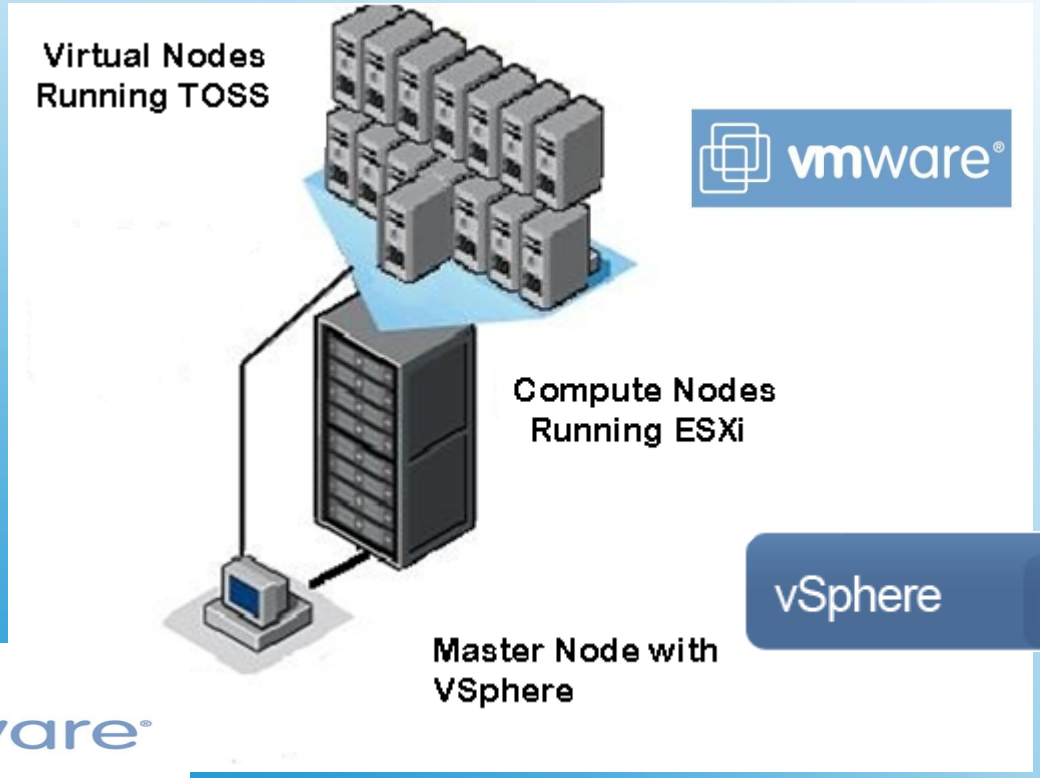
Project

- Tri-lab Operating System Stack (TOSS)
- Slurm – Resource Manager
- Moab – Scheduler
- Gazebo – Acceptance Test Framework
- And more

Challenges

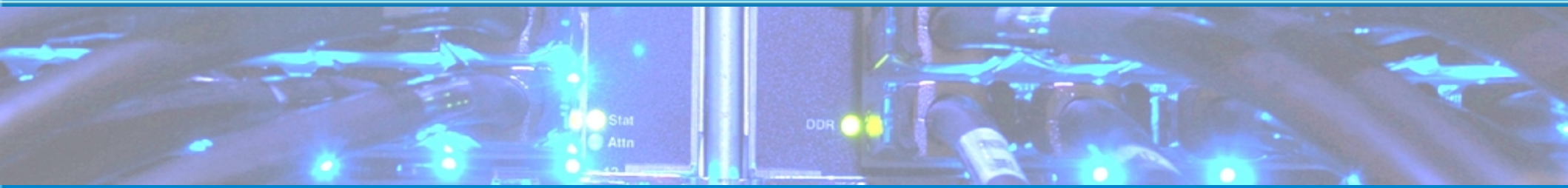
- Virtualization vs Real Hardware
- Management of virtual machines
- Difficulty to set up Gazebo testing

Transition from physical to virtual…



VMware OVF Tool 2.0.1

The OVFTool is a command line utility that gives you the opportunity of importing and exporting OVA packages for installation.

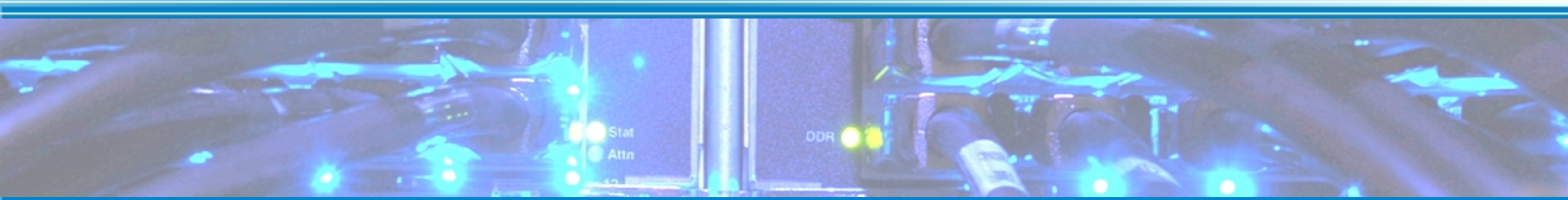


Travis Johnson | Gregorio Hinojos | Benjamin Cheng

What is OVF?

OVF stands for Open Virtualization Format. It contains the metadata information of a virtual machine in XML format. Appealing features of OVF for our project:

- Supports multi virtual configurations
- Enables portable VM packing
- Offers future extensibility



Putting It Together/ Packing It Up

OVA (Open Virtual Appliance). It is a package that contains the files that describe a virtual machine:

- **.OVF file, descriptor file**
- **.MF file, optional manifest**
- **.VMDK file, virtual disk image**

Demo

Want to show:

- The differences of using Infrastructure On Demand vs. VSphere (doing it by hand)
- The Scaling Capability
- Actual User Interface

Benefits of having the Infrastructure On-Demand option

Utilization Improvement

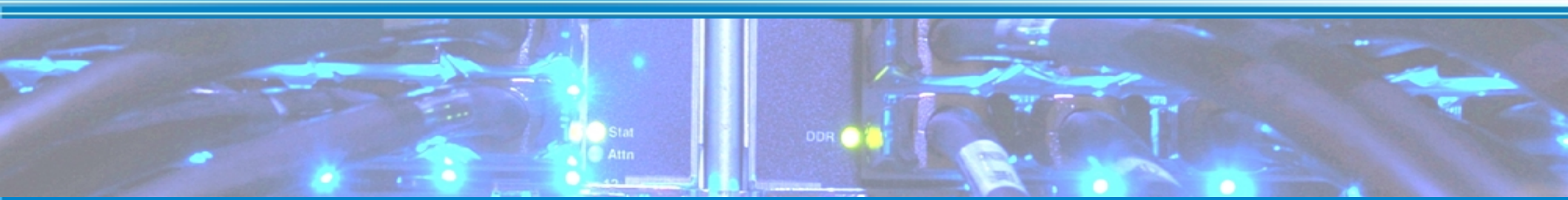
- Reduce load of HPC cluster frontends
- Increased stability of HPC cluster frontends
- Customize time to run jobs
- To offload low priority jobs on to slimly provisioned virtual machines

Scientific Improvement

- Allows more small and large jobs to run
- Legacy code support
- Ability for small jobs such as pre-processing and post-processing virtually
- Customize time to run jobs

Conclusion

- On-Demand Virtualization
  - will not replace conventional clusters
  - alleviates some of the load
  - increases throughput and utilization



Questions?